The Lepskiĭ balancing principle for conjugate gradient regularization

Peter Mathé

Weierstrass Institute Berlin

Warszawa, 09 lutego 2010
Outline

1. Introduction

2. Results
Krylov minimization

Conjugate gradient regularization \((cg)\) is an iterative solver for linear equations in Hilbert space. Here we shall apply this to noisy equations

\[ y^\delta = Tx + \delta \xi, \]

where

- \( T : X \to Y \) has non-closed range (ill-posedness),
- \( \xi \) is bounded deterministic noise,
- \( \delta > 0 \) is the noise level.
Krylov minimization

Conjugate gradient regularization \( \text{(cg) } \) is an iterative solver for linear equations in Hilbert space. Here we shall apply this to noisy equations

\[
y^\delta = Tx + \delta \xi,
\]

where

- \( T : X \rightarrow Y \) has non-closed range (ill-posedness),
- \( \xi \) is bounded deterministic noise,
- \( \delta > 0 \) is the noise level.

\( \text{cg} \) determines at each step \( k \) the element \( x_k^\delta \) as minimizer of

\[
\delta_k := \| y^\delta - Tx_k^\delta \| = \min \left\{ \| y^\delta - Tx \|, \quad x \in \mathcal{K}_k(T^*y^\delta, T^*T) \right\},
\]

where

\[
\mathcal{K}_k(T^*y^\delta, T^*T) = \text{span} \left\{ (T^*T)^{k-1} T^*y^\delta, \quad k = 1, \ldots, k \right\},
\]

denotes the \text{th Krylov-subspace}. 
Background

cg has a clear geometric interpretation, as it determines the descent directions such that the consecutive residuals are $TT^*$-orthogonal, i.e.,

$$\langle T^*(y^\delta - Tx_k^\delta), T^*(y^\delta - Tx_{k+1}^\delta) \rangle = 0.$$ 

Therefore, it should be called conjugate residuals!
The resulting algorithmic description is short, and ’easy to compute’.

Remark

- M. R. Hestenes and E. Stiefel [3], 1952, original study,
- J. R. Shewchuk [5], 1994, “An Introduction to the Conjugate Gradient Method Without the Agonizing Pain“, the most cited introduction to the subject.
Using $cg$ for regularization

In regularization $cg$ is used for the normal equations

\[ T^* y^\delta = T^* Tx + \delta T^* \xi, \]

Since $x_k^\delta$ solves a least-squares problem:
If $y^\delta \not\in \mathcal{R}(T)$ then $\|x_k^\delta\|$ must explode as $k \to \infty$. Therefore regularization (stopping) $k = k(\delta, y^\delta)$ is necessary.

Remark

- A. Nemirowskiĭ [4], 1986, “Regularizing properties of the conjugate gradient method in ill-posed problems”, original study,
- M. Hanke [2], 1995, monograph, polishing the first study,
- Engl, Hanke and Neubauer [1], 1996, chapter of cg, that’s where we refer to!
Challenges

We recall that $x^\delta_k$ solves a minimization problem in a Krylov subspace

$$
x^\delta_k = g_k(T^* T) T^* y^\delta, \quad \deg(g_k) \leq k - 1.
$$

The polynomial $g_k = g_k(y^\delta)$ depends on the data $y^\delta$, therefore $cg$ is a nonlinear iterative method!

Consequently,

- there is no immediate bias-variance decomposition,
- no control of growth of the noise propagation,
- no a priori parameter choice, and
- it is hard to check the qualification of $cg$. 

Challenges

We recall that $x_k^\delta$ solves a minimization problem in a Krylov subspace

$$x_k^\delta = g_k (T^* T) T^* y^\delta, \quad \deg(g_k) \leq k - 1.$$  

The polynomial $g_k = g_k(y^\delta)$ depends on the data $y^\delta$, therefore $cg$ is a **nonlinear iterative method**!

Consequently,

- there is no immediate bias-variance decomposition,
- no control of growth of the noise propagation,
- no a priori parameter choice, and
- it is hard to check the qualification of $cg$.

**Theorem (Nemirovskiǐ)**

*If $x \in \mathcal{R}(T^* T)^\mu$, and if $K_{DP}$ then is according to the discrepancy principle then*

$$\|x - x_{K_{DP}}^\delta\| \leq C \delta^{\mu/(\mu+1/2)}.$$
Goals, achievements

- We want to extend $cg$ to statistical ill-posed problems. Therefore, the discrepancy principle causes problems!
- Need other parameter choice.

In our study we achieved

- application of the *Lepskiǐ balancing principle* for parameter choice,
- to obtain an *oracle inequality*, and
- to extend $cg$ to (a class of) *general smoothness assumptions*. 

**Definition**

An index function $\psi$ is said to be majorized by the power $\mu$ if the function $t \mapsto t^{\mu}/\psi(t)$ is an index function. An index function is said to be majorized by a power if it is majorized by the power $\mu$ for some $\mu > 0$.

**Remark**

Goals, achievements

- We want to extend \( cg \) to statistical ill-posed problems. Therefore, the discrepancy principle causes problems!
- Need other parameter choice.

In our study we achieved

- application of the *Lepskiǐ balancing principle* for parameter choice,
- to obtain an *oracle inequality*, and
- to extend \( cg \) to (a class of) *general smoothness assumptions*.

Definition

An index function \( \psi \) is said to be majorized by the power \( \mu \) if the function \( t \mapsto t^\mu / \psi(t) \) is an index function. An index function is said to be majorized by a power if it is majorized by the power \( \mu \) for some \( \mu > 0 \).
Goals, achievements

- We want to extend \textit{cg} to statistical ill-posed problems. Therefore, the discrepancy principle causes problems!
- Need other parameter choice.

In our study we achieved

- application of the \textit{Lepskiǐ balancing principle} for parameter choice,
- to obtain an \textit{oracle inequality}, and
- to extend \textit{cg} to (a class of) \textit{general smoothness assumptions}.

Definition

An index function $\psi$ is said to be majorized by the power $\mu$ if the function $t \mapsto t^\mu / \psi(t)$ is an index function. An index function is said to be majorized by a power if it is majorized by the power $\mu$ for some $\mu > 0$.

Remark

Outline

1. Introduction

2. Results
Error decomposition

Theorem

Suppose that $\psi$ is majorized by a power and that $x \in H_\psi$. There is a constant $C$ such that

$$
\|x - x_\delta\| \leq C\psi \left( \Theta_\psi^{-1}(d_k) \right) + 3 \frac{\delta}{\sqrt{\alpha_k}},
$$

where $\Theta_\psi(t) = \sqrt{t} \psi(t)$, and

- $d_k := \max \{ \delta, \|y_\delta - Tx_\delta\| \}$, and $\alpha_k := |r_k'(0)|^{-1}$, $k = 1, 2, \ldots$. 

Remark

Notice that $\alpha_k \downarrow 0$ as $k \to \infty$, and $\|y_\delta - Tx_\delta\| \downarrow 0$ as $k \to \infty$. This reflects that the first summand subsumes non-linearity terms. There is no control of increase of $|r_k'(0)|^{-1}$!
Error decomposition

Theorem

Suppose that $\psi$ is majorized by a power and that $x \in H_\psi$. There is a constant $C$ such that

$$
\|x - x^\delta_k\| \leq C\psi \left( \Theta^{-1}_\psi (d_k) \right) + 3 \frac{\delta}{\sqrt{\alpha_k}},
$$

where $\Theta_\psi(t) = \sqrt{t}\psi(t)$, and

- $d_k := \max \{ \delta, \|y^\delta - Tx^\delta_k\| \}$, and $\alpha_k := |r'_k(0)|^{-1}$, $k = 1, 2, \ldots$.

Remark

Notice that

- $\alpha_k \searrow 0$ as $k \to \infty$, and $\|y^\delta - Tx^\delta_k\| \searrow 0$ as $k \to \infty$.
- $d_k \not\to 0$ as $k \to \infty$. This reflects that the first summand subsumes non-linearity terms.
- There is no control of increase of $|r'_k(0)|$!
What is $|r_k'(0)|$?

- Recall that $x_k^\delta := g_k(T^* T) T^* y^\delta$ with polynomial $g_k$.
- We assign $r_k(\lambda) := 1 - \lambda g_k(\lambda)$.
- This is a polynomial of $\deg(r_k) = k$, $r_k(0) = 1$.
- The polynomials $r_k$, $k = 1, \ldots$ are orthogonal with respect to $d\mu(\lambda) := \lambda d\|F_\lambda y^\delta\|^2$.
- The quantity $|r_k'(0)|$ is the (abs. value of) first derivative.
- It can be shown that $|r_k'(0)| \nearrow \infty$ as $k \to \infty$.
- The polynomials are easily calculated along with the steps of cg.
Theorem

Suppose that $\psi$ is majorized by a power, and that $K_{DP}$ is according to the discrepancy principle. Then there is a constant $C < \infty$ such that for $x \in H_\psi$ we have that

$$\|x - x_{K_{DP}}^\delta\| \leq C_\psi \left( \Theta_\psi^{-1}(\delta) \right), \quad \delta \to 0,$$

provided that $K_{DP} > 1$. 
Theorem

Suppose that $\psi$ is majorized by a power, and that $K_{DP}$ is according to the discrepancy principle. Then there is a constant $C < \infty$ such that for $x \in H_\psi$ we have that

$$\|x - x_{K_{DP}}^\delta\| \leq C_\psi \left( \Theta_\psi^{-1}(\delta) \right), \quad \delta \to 0,$$

provided that $K_{DP} > 1$.

- This extends the Nemirovskii result.
- Since qualification can only be checked for classical source conditions, we use the method of distance functions as in B. Hofmann, P. Mathé, Analysis of profile functions for general linear regularization methods, SINUM, 2007.
Oracle inequality

Assume \( \{x_1, \ldots, x_m\} \) is a finite set in a metric space \((M, d)\), and \( x \in M \) ("the truth") satisfies \( d(x, x_i) \leq \frac{1}{2} (\Phi(i) + \Psi(i)), \ i = 1, \ldots, m, \) where \( \Phi : \{1, \ldots, m\} \rightarrow \mathbb{R}^+ \) is non-decreasing, and \( \Psi : \{1, \ldots, m\} \rightarrow \mathbb{R}^+ \) is non-increasing.

Fix any \( K > 1 \), and define the set

\[
\Delta = \{1 \leq j \leq m : d(x_i, x_j) \leq K\Psi(i), \ \text{for all} \ i \leq j\}.
\]
**Oracle inequality**

Assume \( \{x_1, \ldots, x_m\} \) is a finite set in a metric space \((M, d')\), and \( x \in M \) ("the truth") satisfies

\[
d(x, x_i) \leq \frac{1}{2} (\Phi(i) + \Psi(i)), \quad i = 1, \ldots, m,
\]

where \( \Phi : \{1, \ldots, m\} \to \mathbb{R}^+ \) is non-decreasing, and \( \Psi : \{1, \ldots, m\} \to \mathbb{R}^+ \) is non-increasing.

Fix any \( K > 1 \), and define the set

\[
\Delta = \{1 \leq j \leq m : d(x_i, x_j) \leq K\Psi(i), \text{ for all } i \leq j\}.
\]

**Definition**

Any integer satisfying \( \bar{j} \in \Delta \) and \( (\bar{j} + 1) \notin \Delta \) is called Lepskiĭ parameter.
Oracle inequality
Assume \( \{x_1, \ldots, x_m\} \) is a finite set in a metric space \((M, d)\), and \( x \in M \) ("the truth") satisfies \( d(x, x_i) \leq \frac{1}{2} (\Phi(i) + \Psi(i)) \), \( i = 1, \ldots, m \), where \( \Phi : \{1, \ldots, m\} \rightarrow \mathbb{R}^+ \) is non-decreasing, and \( \Psi : \{1, \ldots, m\} \rightarrow \mathbb{R}^+ \) is non-increasing.
Fix any \( K > 1 \), and define the set
\[
\Delta = \{ 1 \leq j \leq m : d(x_i, x_j) \leq K\Psi(i), \text{ for all } i \leq j \}. 
\]

Definition

Any integer satisfying \( \bar{j} \in \Delta \) and \((\bar{j} + 1) \notin \Delta \) is called Lepskiï parameter.

Proposition (Abstract oracle inequality)

Any Lepskiï parameter \( \bar{j} \) satisfies
\[
d(x, x_{\bar{j}}) \leq \frac{1}{2} \min_{1 \leq i \leq m} \left\{ \frac{K}{K - 1} \Phi(i) + (1 + 2K)\Psi(i) \right\}. 
\]
Theorem

Suppose that $\psi$ is majorized by a power, and that $K_L$ is the Lepskiǐ parameter choice. Then there is a constant $C < \infty$ such that for $x \in H_\psi$ we have that

$$\|x - x_{K_L}^\delta\| \leq C_\psi \left( \Theta_\psi^{-1}(\delta) \right), \quad \delta \to 0,$$

provided that $1 < K_L < k_{\text{max}}$. 

**Remark**

Although the oracle bound holds without any assumptions, the optimal order can be guaranteed only if $K_L$ is internal.
Theorem

Suppose that $\psi$ is majorized by a power, and that $K_L$ is the Lepskiĭ parameter choice. Then there is a constant $C < \infty$ such that for $x \in H_\psi$ we have that

$$\| x - x_{K_L}^\delta \| \leq C_\psi \left( \Theta_\psi^{-1}(\delta) \right), \quad \delta \to 0,$$

provided that $1 < K_L < k_{\text{max}}$.

Remark

Although the oracle bound holds without any assumptions, the optimal order can be guaranteed only if $K_L$ is internal.
Heinz W. Engl, Martin Hanke, and Andreas Neubauer.  
*Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*.  

Martin Hanke.  
*Conjugate gradient type methods for ill-posed problems*, volume 327 of *Pitman Research Notes in Mathematics Series*.  

Magnus R. Hestenes and Eduard Stiefel.  
Methods of conjugate gradients for solving linear systems.  

A. S. Nemirovskii.  
Regularizing properties of the conjugate gradient method in ill-posed problems.  

Jonathan R Shewchuk.
An introduction to the conjugate gradient method without the agonizing pain.